

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-295991

(43)Date of publication of application : 10.11.1995

(51)Int.Cl.

G06F 17/28

(21)Application number : 06-082779

(71)Applicant : RICOH CO LTD

(22)Date of filing : 21.04.1994

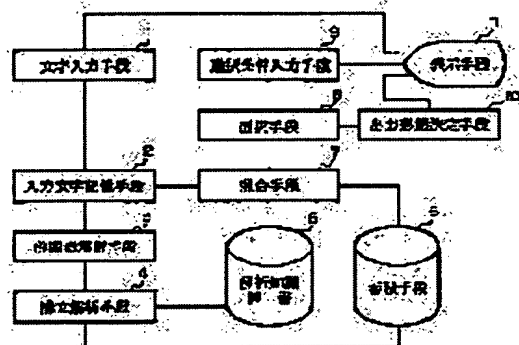
(72)Inventor : OGURO YOSHIHISA
KATOOKA TAKASHI
OONO AYAKO
FUIRISU ANUIRU
NARITA MASUMI

(54) LANGUAGE ANALYZER

(57)Abstract:

PURPOSE: To recognize wording of a vocabulary included in an input sentence by extracting a sentence similar to the input sentence in terms of a language among stored sentences and to obtain a reference of a translated input sentence by extracting a result of a translation of a sentence similar to the input sentence in terms of a language among stored sentences.

CONSTITUTION: The analyzer has a language processing system provided with an entry means 1 to enter a sentence, a storage means 2 storing a sentence, a display means 11, an analysis dictionary storing an intelligence required for language analysis, and a storage means 6 storing sentences stored in the storage means 2. A collation means 7 compares and collates an input sentence with sentences stored in the storage means 6 and a selection condition entry means 9 selects and extracts a sentence in matching with an entered selection condition.



LEGAL STATUS

[Date of request for examination] 01.08.2000

[Date of sending the examiner's decision of rejection] 10.09.2002

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

*** NOTICES ***

Japan Patent Office is not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.*** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] An input means for inputting a sentence, a display means, a storage means to memorize a sentence, an analysis dictionary that has memorized knowledge required for language analysis, It has the language processing system equipped with an are recording means to accumulate a sentence memorized with said storage means. And language analysis equipment characterized by having a selection means to choose the result concerned based on comparison / collating result of a comparison collating means and this comparison collating means which compares and collates an input statement with a sentence accumulated in said are recording means.

[Claim 2] Language analysis equipment according to claim 1 characterized by having an operation condition input means to specify operation conditions of a comparison and collating of said comparison collating means.

[Claim 3] Language analysis equipment according to claim 1 characterized by having an are recording means to accumulate said language analysis result, and comparing and collating an input statement with a language analysis result accumulated in this are recording means.

[Claim 4] Language analysis equipment according to claim 1 characterized by having a language analysis means which carries out language analysis of an input means to input a pair of the 1st language sentence and the 2nd language sentence, a storage means to memorize said a pair of sentence, an are recording means to accumulate said a pair of sentence, said one side of a pair of sentence, or both.

[Claim 5] Language analysis equipment according to claim 4 characterized by having an analysis dictionary for text and its translation between said 2 language.

[Claim 6] Language analysis equipment according to claim 4 characterized by having a synonym dictionary and a semantic thesaurus dictionary.

[Claim 7] Language analysis equipment according to claim 4 characterized by having an exchange means to exchange stored data.

[Translation done.]

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.**** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Industrial Application] This invention relates to details more about language analysis equipment at the language analysis equipment which has the function extracted out of the sentence which had the input statement and the sentence similar to a language target accumulated.

[0002]

[Description of the Prior Art] Conventionally, as a means to search an input statement and a similar sentence, there are Michifumi Sato: "combination use of the example of two or more translations in the translation based on a MBT2:example", Japanese Society for Artificial Intelligence, VOL.6, and No.6 (1991). However, this retrieval technology is offered as a component of the so-called machine translation system of the case base, that result is the thing of the property automatically processed by the system, and a user does not receive it.

[0003]

[Problem(s) to be Solved by the Invention] This invention was made in view of the actual condition like ****, extracts an input statement and a sentence similar to a linguistics target, and is made for the purpose of offering the language analysis equipment which has the function to refer to the information about the sentence.

[0004]

[Means for Solving the Problem] An input means for inputting (1) sentence, in order that this invention may solve the above-mentioned technical problem, A display means, a storage means to memorize a sentence, an analysis dictionary that has memorized knowledge required for language analysis, It has the language processing system equipped with an are recording means to accumulate a sentence memorized with said storage means. and a thing for which it has a selection means to choose the result concerned based on comparison / collating result of a comparison collating means and this comparison collating means which compares and collates an input statement with a sentence accumulated in said are recording means — further (2) — having an operation condition input means to specify operation conditions of a comparison and collating of said comparison collating means — further (3) — having an are recording means to accumulate said language analysis result, and comparing and collating an input statement with a language analysis result accumulated in this are recording means — further (4) It has a language analysis means which carries out language analysis of an input means to input a pair of the 1st language sentence and the 2nd language sentence, a storage means to memorize said a pair of sentence, an are recording means to accumulate said a pair of sentence, said one side of a pair of sentence, or both, (5) — it is further characterized by to have an analysis dictionary for text and its translation between said 2 language, to have (6) synonym dictionaries and a semantic thesaurus dictionary, and having an exchange means to exchange further data by which (7) are recording was carried out. [furthermore,]

[0005]

[Function]

** Get to know the direction for use of the vocabulary contained in an input statement by extracting an input statement and a sentence similar to a language target out of the

accumulated sentence.

**** Obtain reference of a translation of an input statement by extracting the bilingual result of an input statement and a sentence similar to a language target out of the accumulated sentence.**
[0006]

[Example] Drawing 1 is a block diagram as one example of the language-processing equipment which has a similar sentence optional feature by this invention. An alphabetic character input means by which one is represented with a keyboard, OCR equipment, etc. among drawing, An input-statement character storage means by which 2 is represented with a register memory apparatus etc., and 3 A morphological analysis means, As for an are recording means by which an syntax-analysis means and 5 are represented with an analysis knowledge dictionary, and 6 is represented with a memory apparatus, HDD equipment, etc. for 4, a selection-condition input means by which a collating means and 8 are represented with a sentence selection means, and 9 is represented with a keyboard etc. for 7, and 10, an output gestalt decision means and 11 are display means.

[0007] Although first each means to constitute this invention is explained, unless it refuses especially, in the case of a word, in a language pair, English is hereafter explained for a word as the pair of English and Japanese at an example. However, it can be understood easily that they are other language and the thing similarly materialized in a language pair.

[0008] The alphabetic character input means alphabetic character input means 1 inputs the sentence of the 1st language into language analysis equipment by the typing by the keyboard, an OCR input with a scanner, etc. When there is the corresponding 2nd language sentence (equivalent to the bilingual sentence of the 1st language sentence), it inputs collectively.

[0009] The morphological analysis means morphological analysis means 3 divides the inputted sentence into a morpheme (equivalent to a word). In the language which puts in and writes a null with a space between words between words like English, with a conjugation table, it returns to the original form (standard form), a dictionary is consulted from the analysis knowledge dictionary 5, and it asks for an activity, a part of speech, a translation, etc. In the case of the language by which close writing is carried out like Japanese, the optimal word range is called for based on regulations, such as the longest match principle and the number of clauses minimum method. Any of the 1st language sentence and the 2nd language sentence are sufficient as the sentence inputted.

[0010] The syntax-analysis means syntax-analysis means 4 searches for the structure of the inputted sentence according to the syntax accumulated in the analysis knowledge dictionary 5. In syntax, it is an extended context free language (A-CFG).

Extended transition syntax (ATN)

There is a dependency method etc. It is known widely and the method of analyzing a sentence based on these syntax is summarized into standard textbooks (Nagao truth: "mathematical linguistics" ***** 1983 etc.).

[0011] A sentence is collected into every syntax items (a subject, an object, complement, etc.) by the syntax-analysis means 4, and is expressed by the functor tree with it. An easy example is shown in drawing 2 . According to drawing 2 , a sentence "Tom saw Mary with a telescope" is a context free language. S→ NP VP (1)

→ (N) (VP PP) (2)

→ (N) (V NP) (P NP) (3)

→ (N) (V (N)) (P (DET N)) (4)

**** -- it is analyzed like. Setting to regulation application of (3) among this, this sentence is Subj (subject):NVerb(verb):VObj(object):NPP (prepositional phrase) :P It turns out that it is a sentence type called Subj+Verb+Obj+PP NP since there is becoming correspondence, and the functor-role of each word also becomes clear.**

[0012] The are recording means are recording means 6 accumulates the sentence inputted by the alphabetic character input means in large quantities. Any of the 1st language sentence and the 2nd language sentence are sufficient as a sentence. Both are memorized maintaining correspondence relation, when there is the 2nd language sentence (equivalent to the bilingual sentence of the 1st language sentence) corresponding to the 1st language sentence.

Furthermore, the sentence analyzed by the morphological analysis means 8 and the syntax-analysis means 9 is also accumulated.

[0013] The collating processing collating means 4 compares the thing of the same syntax item in two sentences (that as which the user inputted one side into the input-statement character storage means 2, thing by which another side was accumulated in the are recording means 6) collated.

[0014] If the selection-condition input means selection-condition input means 9 follows the conditions for which a user asks, it will be chosen, and the result of having performed collating with an input statement and the accumulated sentence with the collating means 7 will not choose it, if it does not follow. It is selection etc., if it is the sentence of a S+V+O+O mold, a selection subject is not a pronoun, a selection object is a character string "a book" and the selected character train "a book" is included as conditions which a user inputs, for example. The selection condition which can be performed can be beforehand created and set by menu-driven etc., when the dictionary and the syntax-analysis method of using it are decided, since it is dependent on the contents of a dictionary, or the syntax-analysis method.

[0015] A user directs the output gestalt decision means output gestalt decision means 10 so that the sentence chosen by the selection means 8 can be outputted with a desired gestalt. For example, they are the character string of the sentence as which only the character string of the selected sentence was chosen, the character string of the sentence by which analysis result selection was made, an analysis result, the character string of the sentence by which collating result selection was made, the 2nd language sentence that corresponds (if it is). Since the output gestalt which can be performed is known, it creates by menu-driven etc. beforehand and a user should just choose from there.

[0016] Example 1 (it corresponds to claims 1 and 2)

The sentence inputted by the alphabetic character input means 1 is memorized by the input-statement character storage means 2, is compared and collated with the sentence accumulated in the are recording means 6 by the collating means 7, and extracts a similar sentence. It will choose, if the conditions for which the user inputted asks with the selection-condition input means 9 are followed, and the result of a comparison and collating will not choose, if it does not follow. In addition, the conditions which a user inputs in this case are as the term of the above-mentioned selection-condition input means having described.

[0017] Drawing 3 is the flow chart of the above-mentioned similar sentence extract processing. step1: Extract a sentence similar to the sentence inputted here out of the accumulated sentence. As an input means, input units, such as keyboard equipment and OCR equipment, are equivalent to this, for example.

step2: Analyze the sentence inputted by step1 with an analysis means.

step3: It extracts one sentence at a time in order out of the sentence accumulated by the are recording means. step4: Processing will be ended if the sentence accumulated by the are recording means is exhausted. If one sentence is able to be extracted, it will progress to step5.

step5: Analyze the sentence extracted by step3 with an analysis means.

step6: Collate the analysis result of the input statement obtained by step2, and the analysis result of the are recording sentence obtained by step3 with a collating means.

step7: If the collating result obtained by step6 is filling the selection condition which the user inputted, it will progress to step8.

step8: If there are a sentence chosen by step7 and hope, the analysis result will be outputted to a display means. It returns to step4.

[0018] Example 2 (it corresponds to claim 3)

Syntax-analysis processing takes time amount in many cases, in order to perform comparatively complicated processing. Whenever an input statement is inputted, if syntax-analysis processing is performed, great time amount is taken for a user to get a similar sentence, and a feeling of use is spoiled remarkably each time. Then, syntax analysis of the inputted sentence or the accumulated sentence is carried out beforehand, and the result is accumulated with the sentence. When carrying out like this and a user inputs a desired sentence, it is not necessary to carry out only syntax analysis of an input statement and to analyze the syntax of an are

recording sentence each time that what is necessary is just to perform collating processing to the analysis result of an are recording sentence. Therefore, since a user can get a sentence similar to a desired sentence promptly, a feeling of use is improved remarkably.

[0019] The flow chart of this processing is shown in drawing 4 (a).

step1: It extracts one sentence at a time in order out of the sentence accumulated by the are recording means. step2: Processing will be ended if the sentence accumulated by the are recording means is exhausted. If one sentence is able to be extracted, it will progress to step5. step3: Analyze the sentence extracted by step1 with an analysis means.

step4: Accumulate the analysis result obtained by step3 with an are recording means. It returns to step1. step5: Extract a sentence similar to the sentence inputted here out of the accumulated sentence. As an input means, input units, such as keyboard equipment and OCR equipment, are equivalent to this, for example.

step6: Processing will be ended if an input statement is exhausted. If there is an input statement, it will progress to step7.

step7: Analyze the sentence inputted by step6.

step8: The analysis result obtained by step7 and the analysis results of all the accumulated sentences are collated, and if a selection condition is satisfied, it will choose (a detailed flow is referring to drawing 5). It returns to step5.

[0020] Drawing 5 collates the analysis result of an input statement, and the analysis results of all the accumulated sentences, and is a flow of processing to choose.

step 8-1: Extract the result for one sentence from the analysis result of an are recording sentence.

step 8-2: If the analysis result of an are recording sentence is exhausted, it will end. If it is able to extract by one sentence, it will progress to step 8-3.

step 8-3: Collate the analysis result of an input statement, and the analysis result of an are recording sentence with a collating means.

If the collating result obtained by step8-4:step 8-4 fills a selection condition, it will progress to step 8-5. If it does not fill, it will return to step 8-1.

If there are a sentence chosen by step8-5:step 8-4 and hope, the analysis result will be outputted to a display means. It returns to step 8-1.

[0021] Example 3 (it corresponds to claim 4)

When translating the 1st language into the 2nd language, there are few possibilities that a right decodement will be obtained, only by transposing the word in the 1st language sentence to the 2nd language word in a bilingual dictionary. Generally, one word has two or more semantics, and semantics is decided from relation with other words in many cases. Therefore, in order to translate correctly, it is necessary to refer to the example of a translation of the past translated correctly. However, when the example which is applied to a desired sentence exactly and which was translated correctly exists, it is few, and in almost all cases, the example of a translation of a desired sentence (or a part of sentence) and a similar sentence is translated into reference.

[0022] About obtaining a sentence similar to a desired sentence (or a part of sentence), it is realizable with invention which made reference in said examples 1 and 2 (claims 1-3). Therefore, what is necessary is to accumulate the pair (namely, example of a translation) of not only a single language sentence but the 1st language sentence (text), and the 2nd language sentence (decodement), and just to choose as coincidence not only the 1st language sentence but the 2nd language sentence which is the pair in invention which made reference in the examples 1 and 2 (claims 1-3), in case the 1st language sentence similar to a desired sentence is extracted.

[0023] For example, suppose that text and its translation are managed as shown in a table 1 with an are recording means. invention which made reference in the examples 1 and 3 (claims 1-3) — English-Japanese — to one of sentences, collating processing is carried out and it determines whether choose a bilingual pair based on the result. in addition, the table 1 — setting — a pair — when description has only either like 3, it functions as invention which made reference in the examples 1 and 2 (claims 1-3) similarly.

[0024]

[A table 1]

対訳を管理する例

対 1	英	Excess adhesive shall be carefully removed to prevent running to other surfaces.
	日	余分な接着剤が他の面に流れていかにように、注意深く除去する。
対 2	英	Extend wing flaps and remove access covers from top leading edge of wing flap.
	日	主翼フラップを下げて、主翼フラップの前縁上面から点検カバーを取り外す。
対 3	英	Remove access covers from bottom of right wing below right drive pulley.
	日	
		*** THE END ***

[0025] Example 4 (it corresponds to claim 5)

In case the similar example of a translation is searched, in some cases, I want to search the sentence of the 1st language with the 2nd language. For example, what is necessary is just to carry out invention which made reference in the examples 1 and 2 (claims 1-3) to the sentence by the side of the 2nd language (Japanese), if the bilingual pair is accumulated to extract the 1st language sentence containing the word of the semantics of "removing" in the 2nd language as shown in a table 1 although the word of the 1st language itself does not specify. However, it is rare that a bilingual pair can obtain in large quantities, and the cost concerning acquisition also tends to become great and it is difficult to accumulate a bilingual pair in large quantities therefore in many cases.

[0026] Then, what is necessary is just to carry out invention which made reference in the example 2 (claim 3) to the 2nd language sentence (or a part of sentence) which changed the 1st language sentence into the 2nd language sentence automatically, and changed it using the bilingual dictionary between the 1st language sentence and the 2nd language sentence. Here, the machine-translation processing itself [so-called] to which current commercialization is performed is sufficient as the automatic conversion in the 2nd language sentence from the 1st language sentence, and it functions enough also by replacement of the simple word which makes reference below.

[0027] The case where English containing the alphabetic word which may become a Japanese reason for "Removing" is extracted out of the sentence in which the contents of the table 1 were accumulated is explained. A bilingual (EIWA) dictionary like drawing 6 is prepared. [] Inside expresses a part-of-speech mark and ";" expresses a pause of a translation. English of a table 1 is replaced by the 2nd language for every word using a bilingual dictionary. for example, a pair -- in the case of English "Excess adhesive shall be carefully removed --" of 1, it becomes as it is shown in a table 2.

[0028]

[A table 2]

第1言語から第2言語に単語毎に置換した例

第1言語	第2言語情報
Excess	〔形〕 超過した； 余分な 〔名〕 超過； 過剰； 過度
adhesive	〔形〕 粘着性の 〔名〕 接着剤； 粘着物
shall	〔助〕 ～だろう； ～させる； ～すべきである
be	〔動〕 ～である； ～がある 〔助〕 ～される； ～している
carefully	〔副〕 注意深く； 慎重に； 入念に
removed	〔動〕 移す； 除去〔移動〕する； 解任する 〔名〕 移動； 移転
***	***
***	***
***	***

[0029] if a desired word "it removes" is searched to the 2nd language information column of a

table 2 — removed ; to [**] move —; [name] migration; to remove [migration] — it turns out that a move agrees. [to dismiss] Therefore, this English is chosen. It carries out to all the sentences that are having the same processing accumulated.

[0030] Example 5 (it corresponds to claim 6)

Reference was made by the example 1 thru/or 4 (claims 1-5) in the method of choosing a sentence similar in surface. In order to investigate the direction for use of a word, it is necessary to refer not only to a sentence similar in surface but to a semantically similar sentence. By comparing the direction for use of a semantically similar word, it is because the feature of direction for use clarifies further. What is necessary is just to use a synonym dictionary and a semantic thesaurus dictionary, in order to extract out of the sentence which is having the semantically similar sentence accumulated. What is necessary is to consider that the word in the same term of a synonym dictionary is the same, and just to carry out invention which made reference in the examples 1-4 (claims 1-5). For example, when the contents of the synonym dictionary are what is shown in a table 3, the word in the synonym column considers that it is the same over processings at large, such as retrieval, collating, and selection.

[0031]

[A table 3]

同義語辞書の例

意味	同義語
取り除くの意	clear, filter, free, leave, reform, remove, relieve, sweep, take

[0032] By using a synonym dictionary, semantics was able to choose the similar sentence about the same word, i.e., a paraphrasing phrase. However, there is not a paraphrasing phrase but a case where he wants to investigate, about the phrase of a superordinate concept or a subordinate concept, and a contrary concept semantically. For example, the feature of direction for use clarifies further by extracting the sentence of the "saving mind" of reverse and comparing both with the "removing mind" of a previous example. For that purpose, what is necessary is just to have a semantic thesaurus dictionary as well as a synonym dictionary. As a semantic thesaurus dictionary, the "classification lexical table" (National Japanese Language Research Institute) and the "class word new lexicon" (Kadokawa Shoten Publishing) are known for the Japanese thing. Moreover, a large number [in the thing of a foreign language / Roget's Thesaurus (P. M.Roget: "Thesaurus of English Words and Phrases", London 1851) is begun, and].

[0033] As the implementation method, the system of a semantic marker as shown in drawing 7 - 9 is created, for example, and the semantic marker is given to each keyword of an analysis dictionary. A user inputs [a high order or low order, and] how many it considers that the hierarchy of less than is the same. What is necessary is just to judge with having agreed, when the relation of the semantic marker of the word in an input statement and the word in an are recording sentence was less than the number of hierarchies that the user inputted in the case of collating processing.

[0034] Example 5 (it corresponds to claim 7)

the time amount which retrieval will take if the amount of the sentence accumulated increases - - long - - becoming - - a feeling of use - - ***** - - disadvantage - - it is divided. What is necessary is just to lessen the sentence for retrieval, in order to shorten retrieval time. The sentence accumulated by the are recording means is specifically classified, it is made the bundle, and a user enables it to direct a retrieval range. Since to clarify a source is desired especially in language investigation, if it classifies for every source, retrieval time not only becomes short, but it can obtain the result which is linguistically meaningful. Although it is written with the block diagram of drawing 10 that two or more are recording means 6-1, 6-2, and 6-3 stand in a row, the gestalt will not be asked, if the are recording means does not need to be classified physically and a retrieval range can be specified logically.

[0035] This invention is not limited to the above example and added to a means to accumulate

an input statement in large quantities, the means which compares and collates two or more language analysis results, a means to choose the result concerned based on comparison / collating result, and a means to specify the operation conditions of a comparison and collating. A means to input the pair of a means to accumulate a language analysis result, the 1st language sentence, and the 2nd language sentence, It is characterized by having a means to memorize the pair, a means to accumulate the pair in large quantities, the means of the pair that reaches on the other hand and carries out language analysis of both, the analysis dictionary for text and its translation between 2 language, a synonym dictionary, and a semantic thesaurus dictionary. By this (1) by extracting an input statement and a sentence similar to a language target out of the accumulated sentence Essence is to offer language analysis equipment equipped with the extracting-out of accumulated sentence-bilingual result of function [to get to know the direction for use of the vocabulary contained in an input statement], and (2) input-statements and sentence similar to language target function.

[0036]

[Effect of the Invention] In invention of claims 1 and 2, the direction for use of the vocabulary contained in an input statement can be known by extracting out of the sentence which had the input statement and the sentence similar to a language target accumulated. In invention of claim 3, that what is necessary is just to perform collating processing to the analysis result of an are recording sentence, since it is not necessary to analyze the syntax of an are recording sentence, and a user can get a sentence similar to a desired sentence promptly, a feeling of use is improved remarkably each time. In invention of claim 4, since the example of a translation of a sentence similar to a desired sentence can be acquired, it becomes the reference at the time of translating a desired sentence, and a translation can be carried out to high degree of accuracy. In invention of claim 5, since the 1st language sentence similar to the 2nd language sentence can be searched, the example of a translation can be acquired without needing bilingual data. In invention of claim 6, since a similar or opposite sentence can be searched semantically, the feature of direction for use can be investigated more clearly. In invention of claim 7, since a user can specify a retrieval range, retrieval time is shortened, and since a feeling of use is not only improved remarkably, but it can classify and keep an are recording sentence, it becomes possible to clarify a source.

[Translation done.]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-295991

(43) 公開日 平成7年(1995)11月10日

(51) IntCl. ⁶	識別記号	序内整理番号	F I	技術表示箇所
G 0 6 F 17/28		8219-5L	G 0 6 F 15/ 38	Z
		8219-5L		C

審査請求 未請求 請求項の数7 O L (全 9 頁)

(21) 出願番号 特願平6-82779

(22) 出願日 平成6年(1994)4月21日

(71) 出願人 000006747

株式会社リコー

東京都大田区中馬込1丁目3番6号

(72) 発明者 大黒 慶久

東京都大田区中馬込1丁目3番6号 株式会社リコー内

(72) 発明者 加登岡 隆

東京都大田区中馬込1丁目3番6号 株式会社リコー内

(72) 発明者 大野 亜矢子

東京都大田区中馬込1丁目3番6号 株式会社リコー内

(74) 代理人 弁理士 高野 明近

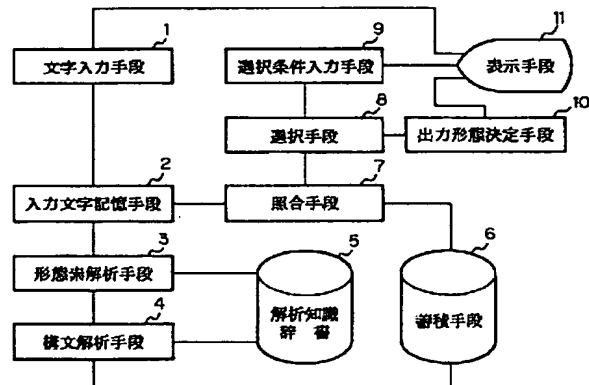
最終頁に続く

(54) 【発明の名称】 言語解析装置

(57) 【要約】

【目的】 入力文と言語的に類似する文を、蓄積された文の中から抽出することによって、入力文中に含まれる語彙の用法を知る。入力文と言語的に類似する文の対訳結果を、蓄積された文の中から抽出することによって、入力文の翻訳の参照を得る。

【構成】 文を入力するための入力手段1、文を記憶する記憶手段2、表示手段11、言語解析に必要な知識を記憶している解析辞書5、前記記憶手段2で記憶した文を蓄積する蓄積手段6を備えた言語処理系を有する。照合手段7により、入力文を蓄積手段6に蓄積されている文と比較・照合し、選択条件入力手段9により入力される選択条件に合った文を選択、抽出する。



本発明の原理を示すブロック図

【特許請求の範囲】

【請求項 1】 文を入力するための入力手段、表示手段、文を記憶する記憶手段、言語解析に必要な知識を記憶している解析辞書、前記記憶手段で記憶した文を蓄積する蓄積手段を備えた言語処理系を有し、かつ、入力文を前記蓄積手段に蓄積された文と比較・照合する比較照合手段、該比較照合手段の比較・照合結果に基づいて当該結果を選択する選択手段を有することを特徴とする言語解析装置。

【請求項 2】 前記比較照合手段の比較・照合の実施条件を指定する実施条件入力手段を有することを特徴とする請求項 1 に記載の言語解析装置。

【請求項 3】 前記言語解析結果を蓄積する蓄積手段を有し、入力文を該蓄積手段に蓄積された言語解析結果と比較・照合することを特徴とする請求項 1 に記載の言語解析装置。

【請求項 4】 第 1 言語文および第 2 言語文の対を入力する入力手段、前記対の文を記憶する記憶手段、前記対の文を蓄積する蓄積手段、前記対の文の一方又は両方を言語解析する言語解析手段を有することを特徴とする請求項 1 に記載の言語解析装置。

【請求項 5】 前記 2 言語間の対訳用解析辞書を有することを特徴とする請求項 4 に記載の言語解析装置。

【請求項 6】 同義語辞書および意味シソーラス辞書を有することを特徴とする請求項 4 に記載の言語解析装置。

【請求項 7】 蓄積されたデータを交換する交換手段を有することを特徴とする請求項 4 に記載の言語解析装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、言語解析装置に関し、より詳細には、入力文と言語的に類似な文を蓄積された文の中から抽出する機能を有する言語解析装置に関する。

【0002】

【従来の技術】 従来、入力文と類似する文を検索する手段としては、佐藤理史：“MBT2：実例に基づく翻訳における複数翻訳例の組合せ利用”、人工知能学会誌、VOL.6, No.6 (1991) がある。しかし、この検索技術は、いわゆる事例ベースの機械翻訳システムの構成要素として供されたものであり、その結果はシステムによって自動的に処理される性質のものであって、使用者が受け取るものではない。

【0003】

【発明が解決しようとする課題】 本発明は、上述のごとき実情に鑑みてなされたもので、入力文と言語的に類似な文を抽出し、その文に関する情報を参照する機能を有する言語解析装置を提供することを目的としてなされたものである。

【0004】

【課題を解決するための手段】 本発明は、上記課題を解決するために、(1) 文を入力するための入力手段、表示手段、文を記憶する記憶手段、言語解析に必要な知識を記憶している解析辞書、前記記憶手段で記憶した文を蓄積する蓄積手段を備えた言語処理系を有し、かつ、入力文を前記蓄積手段に蓄積された文と比較・照合する比較照合手段、該比較照合手段の比較・照合結果に基づいて当該結果を選択する選択手段を有すること、更には、

(2) 前記比較照合手段の比較・照合の実施条件を指定する実施条件入力手段を有すること、更には、(3) 前記言語解析結果を蓄積する蓄積手段を有し、入力文を該蓄積手段に蓄積された言語解析結果と比較・照合すること、更には、(4) 第 1 言語文および第 2 言語文の対を入力する入力手段、前記対の文を記憶する記憶手段、前記対の文を蓄積する蓄積手段、前記対の文の一方又は両方を言語解析する言語解析手段を有すること、更には、

(5) 前記 2 言語間の対訳用解析辞書を有すること、更には、(6) 同義語辞書および意味シソーラス辞書を有すること、更には、(7) 蓄積されたデータを交換する交換手段を有することを特徴としたものである。

【0005】

【作用】

① 入力文と言語的に類似する文を、蓄積された文の中から抽出することによって、入力文中に含まれる語彙の用法を知る。

② 入力文と言語的に類似する文の対訳結果を、蓄積された文の中から抽出することによって、入力文の翻訳の参照を得る。

【0006】

【実施例】 図 1 は、本発明による類似文選択機能を有する言語処理装置の一実施例としてのブロック図で、図中、1 はキーボード、OCR 装置などで代表される文字入力手段、2 はレジスタ・メモリ装置などで代表される入力文字記憶手段、3 は形態素解析手段、4 は構文解析手段、5 は解析知識辞書、6 はメモリ装置、HDD 装置などで代表される蓄積手段、7 は照合手段、8 は文選択手段、9 はキーボードなどで代表される選択条件入力手段、10 は出力形態決定手段、11 は表示手段である。

【0007】 最初に、本発明を構成する各手段について説明するが、以下、特にことわらない限り、一言語の場合は英語を、言語対の場合には英語と日本語との対を例に説明する。しかし、他の言語、および言語対においても同様に成立するものであることは容易に理解できよう。

【0008】 文字入力手段

文字入力手段 1 は、第 1 言語の文を、キーボードによるタイプ入力、スキャナによる OCR 入力などによって言語解析装置に入力する。対応する第 2 言語文 (第 1 言語文の対訳文に相当) がある場合には、併せて入力する。

【0009】形態素解析手段

形態素解析手段3は、入力された文を、形態素（単語に相当）に分割する。英語のように単語と単語との間に空白をいれて分ち書きする言語では、活用変化表によって原形（標準形）に戻し、解析知識辞書5より辞書引きして、活用、品詞、訳語などを求める。日本語のようにべた書きされる言語の場合、最長一致法、文節数最小法などの規則に基づいて、最適な単語範囲が求められる。入力される文は第1言語文、第2言語文のいずれでも構わない。

【0010】構文解析手段

構文解析手段4は、入力された文の構造を、解析知識辞書5に蓄積されている文法にしたがって求める。文法に

$$\begin{aligned} S &\rightarrow NP \quad VP & (1) \\ &\rightarrow (N) (VP \quad PP) & (2) \\ &\rightarrow (N) ((V \quad NP) (P \quad NP)) & (3) \\ &\rightarrow (N) ((V (N)) (P (DET \quad N))) & (4) \end{aligned}$$

のように解析されている。この内、(3)の規則適用において、この文は

Subj（主語）：N

Verb（動詞）：V

Obj（目的語）：N

PP（前置詞句）：P NP

なる対応があるので、

Subj+Verb+Obj+PP

という文型であるとわかり、各単語の構文的な役割も明らかになる。

【0012】蓄積手段

蓄積手段6は、文字入力手段によって入力された文を大量に蓄積する。文は第1言語文、第2言語文のいずれでもよい。第1言語文に対応する第2言語文（第1言語文の対訳文に相当）がある場合には対応関係を保ったまま、両者を記憶する。更には、形態素解析手段8、構文解析手段9によって解析された文をも蓄積する。

【0013】照合処理

照合手段4は、照合される二つの文（一方は使用者が入力文字記憶手段2に入力したもの、もう一方は蓄積手段6に蓄積されたもの）において同じ構文要素のものを比較する。

【0014】選択条件入力手段

選択条件入力手段9は、照合手段7によって入力文と蓄積された文との照合を行った結果が、使用者の所望する条件に従うならば選択し、従わないならば選択しない。

使用者の入力する条件としては、例えば、

S+V+O+O型の文なら選択

主語が代名詞でないなら選択

目的語が文字列「本」なら選択

文字列「本」を含んでいるなら選択

などである。実行可能な選択条件は、辞書内容や構文解析方法に依存するので、使用する辞書や構文解析方法が

は、

拡張文脈自由文法（A-CFG）

拡張遷移文法（ATN）

係受け法

などがある。これらの文法に基づいて文を解析する方法については、広く知られており、標準的な教科書（長尾真：「言語工学」昭光堂1983など）にまとめられている。

【0011】構文解析手段4によって、文は、構文要素（主語、目的語、補語など）毎にまとめあげられ、構文木によって表現される。簡単な例を図2に示す。図2によると、文“Tom saw Mary with a telescope”は、文脈自由文法によって、

決まった時点で、予めメニュー方式などで作成しておく。

【0015】出力形態決定手段

出力形態決定手段10は、選択手段8によって選択された文を所望の形態で出力できるよう、使用者が指示する。例えば、

選択された文の文字列のみ

選択された文の文字列、解析結果

選択された文の文字列、解析結果、照合結果

選択された文の文字列、（あれば）対応する第2言語文などである。実行可能な出力形態は既知であるから、予めメニュー方式などで作成しておき、使用者はそこから選択すればよい。

【0016】実施例1（請求項1及び2に対応）

文字入力手段1によって入力された文は、入力文字記憶手段2に記憶され、照合手段7によって、蓄積手段6に蓄積されている文と比較・照合して類似文を抽出する。比較・照合の結果が、選択条件入力手段9によって、入力される使用者の所望する条件に従うならば選択し、従わないならば選択しない。なお、この場合、使用者が入力する条件は、前述の選択条件入力手段の項で述べた通りである。

【0017】図3は、前述の類似文抽出処理のフローチャートである。

step1：ここで入力された文に類似する文を、蓄積された文の中から抽出する。入力手段としては、例えば、キーボード装置やOCR装置などの入力装置がこれに相当する。

step2：step1で入力された文を解析手段によって解析する。

step3：蓄積手段によって蓄積されている文の中から、一文ずつ順に抽出する。

step4：蓄積手段によって蓄積されている文が尽きた

ら、処理を終了する。一文を抽出できたならば、step 5へすすむ。

step 5 : step 3 で抽出された文を解析手段によって解析する。

step 6 : step 2 で得た入力文の解析結果とstep 3 で得た蓄積文の解析結果とを照合手段によって照合する。

step 7 : step 6 で得た照合結果が、使用者が入力した選択条件を満たしているならば、step 8 にすすむ。

step 8 : step 7 で選択した文、および希望があればその解析結果を表示手段に出力する。step 4 に戻る。

【0018】実施例 2 (請求項 3 に対応)

構文解析処理は、比較的複雑な処理を行なうために時間がかかる場合が多い。入力文が入力される毎にその都度、構文解析処理を行なっているのは、使用者が類似文を得るまで多大な時間を要し、使用感を著しく損なう。そこで、入力された文、或いは蓄積された文の構文解析を予め実施し、その結果を文とともに蓄積しておく。こうすれば、使用者が所望の文を入力した際に、入力文の構文解析のみを実施し、蓄積文の解析結果に対して照合処理を行なうだけでよく、その都度、蓄積文を構文解析する必要はない。よって、使用者は、所望の文に類似する文を速やかに得ることができるので、使用感が著しく改善される。

【0019】図 4 (a) に本処理のフローチャートを示す。

step 1 : 蓄積手段によって蓄積されている文の中から、一文ずつ順に抽出する。

step 2 : 蓄積手段によって蓄積されている文が尽きたら、処理を終了する。一文を抽出できたならば、step 5へすすむ。

step 3 : step 1 で抽出された文を解析手段によって解析する。

step 4 : step 3 で得た解析結果を蓄積手段によって、蓄積する。step 1 に戻る。

step 5 : ここで入力された文に類似する文を、蓄積された文の中から抽出する。入力手段としては、例えば、キーボード装置やOCR装置などの入力装置がこれに相当する。

step 6 : 入力文が尽きたら処理を終了する。入力文があるならば、step 7へすすむ。

step 7 : step 6 で入力された文を解析する。

step 8 : step 7 で得た解析結果と、蓄積された文すべての解析結果とを照合し、選択条件を満足すれば選択する(詳細なフローは、図 5 参照)。step 5 に戻る。

【0020】図 5 は、入力文の解析結果と、蓄積された

文すべての解析結果とを照合し、選択する処理のフローである。

step 8-1 : 蓄積文の解析結果から一文分の結果を抽出する。

step 8-2 : 蓄積文の解析結果が尽きたならば終了する。一文分抽出できたならば、step 8-3 にすすむ。

step 8-3 : 入力文の解析結果と蓄積文の解析結果とを照合手段により照合する。

step 8-4 : step 8-4 で得た照合結果が選択条件を満たすならば、step 8-5 にすすむ。満たさないならば、step 8-1 に戻る。

step 8-5 : step 8-4 で選択した文、および希望があればその解析結果を表示手段に出力する。step 8-1 に戻る。

【0021】実施例 3 (請求項 4 に対応)

第 1 言語を第 2 言語に翻訳する場合、第 1 言語文中の単語を単に対訳辞書によって第 2 言語単語に置き換えるだけでは、正しい翻訳文が得られる可能性は少ない。一般的に、一つの単語は複数の意味を持ち、他の単語との関係から意味が確定される場合が多い。したがって、正しく翻訳するには、正しく翻訳された過去の翻訳例を参考にする必要がある。しかしながら、所望の文にちょうどあてはまる正しく翻訳された例が存在する場合は少なく、ほとんどの場合、所望の文(あるいは文の一部)と類似する文の翻訳例を参考に翻訳する。

【0022】所望の文(あるいは文の一部)と類似な文を得ることについては、前記実施例 1 及び 2 (請求項 1~3) で言及した発明によって実現できる。したがって、実施例 1 及び 2 (請求項 1~3) で言及した発明において、単一の言語文だけでなく、第 1 言語文(原文)と第 2 言語文(翻訳文)との対(すなわち翻訳例)を蓄積しておき、所望の文に類似する第 1 言語文を抽出する際、第 1 言語文だけでなく、その対である第 2 言語文も同時に選択すればよい。

【0023】例えば、蓄積手段によって、表 1 のように対訳が管理されているとする。実施例 1 及び 3 (請求項 1~3) で言及した発明によって、英日どちらか一方の文に対して、照合処理を実施し、その結果に基づいて対訳対を選択するか否かを決定する。なお、表 1 において、対 3 のようにどちらか一方しか記述がない場合には実施例 1 及び 2 (請求項 1~3) で言及した発明と同じように機能する。

【0024】

【表 1】

対訳を管理する例

対	英	Excess adhesive shall be carefully removed to prevent running to other surfaces.
	日	余分な接着剤が他の面に流れていかにように、注意深く除去する。
対	英	Extend wing flaps and remove access covers from top leading edge of wing flap.
	日	主翼フラップを下げて、主翼フラップの前縁上面から点検カバーを取り外す。
対	英	Remove access covers from bottom of right wing below right drive pulley.
	日	
		...

【0025】実施例4（請求項5に対応）

類似な翻訳例を検索する際、第2言語で第1言語の文を検索したいことがある。例えば、第1言語の単語そのものは特定しないが、第2言語において「除去する」という意味の単語を含む第1言語文を抽出したい場合、表1

20

のように対訳対が蓄積されていれば、第2言語（日本語）側の文に対して、実施例1及び2（請求項1～3）で言及した発明を実施すればよい。しかしながら、対訳対が大量に入手できることは希であり、入手にかかるコストも多大になりがちであり、よって対訳対を大量に蓄積することは困難な場合が多い。

【0026】そこで、第1言語文と第2言語文との間の対訳辞書を用いて、第1言語文を第2言語文に自動的に変換し、変換した第2言語文（あるいは文の一部）に対して、実施例2（請求項3）で言及した発明を実施すれ

30

ばよい。ここで、第1言語文から第2言語文への自動的な変換とは、現在商用化が行なわれている、いわゆる機械翻訳処理そのものでも構わないし、以下で言及する単純な単語の置き換えでも十分機能する。

【0027】「除去する」という日本語訳になる可能性のある英単語を含む英文を、表1の内容の蓄積された文の中から抽出する場合について説明する。図6のような対訳（英和）辞書を用意する。〔 〕内は品詞記号、“；”は訳語の句切りを表わす。表1の英文を対訳辞書を用いて単語毎に第2言語に置換する。例えば、対1の英文“Excess adhesive shall be carefully removed …”の場合、表2のようになる。

【0028】

【表2】

第1言語	第2言語情報
Excess	〔形〕 超過した；余分な 〔名〕 超過；過剰；過度
adhesive	〔形〕 粘着性の 〔名〕 接着剤；粘着物
shall	〔助〕 ～だろう；～させる；～すべきである
be	〔助〕 ～である；～がある 〔助〕 ～される；～している
carefully	〔副〕 注意深く；慎重に；入念に
removed	〔動〕 移す；除去〔移動〕する；解任する 〔名〕 移動；移転
...	...
...	...
...	...

【0029】表2の第2言語情報欄に対して、所望の単語「除去する」を検索すると

40

removed 〔動〕 移す；除去〔移動〕する；解任する

〔名〕 移動；移転

が合致することがわかる。したがって、この英文は選択される。同様な処理を蓄積されている全ての文に対して実施する。

【0030】実施例5（請求項6に対応）

表層的に類似な文を選択する方法を実施例1乃至4（請求項1～5）で言及した。単語の用法を調べるためには、表層的に類似な文だけでなく、意味的に類似する文

50

を参考にする必要がある。意味的に類似する単語の用法を比較することによって、用法の特徴がさらにはっきりしてくるからである。意味的に類似する文を蓄積されている文の中から抽出するには、同義語辞書および意味ソーラス辞書を利用すればよい。同義語辞書の同じ項にある単語は同一とみなして、実施例1～4（請求項1～5）で言及した発明を実施すればよい。例えば、同義語辞書の内容が表3に示すものである場合、同義語欄にある単語は、検索・照合・選択など処理全般にわたって同一とみなす。

【0031】

【表 3】

同義語辞書の例

意味	同義語
取り除くの意	clear, filter, free, leave, reform, remove, relieve, sweep, take

【0032】同義語辞書を用いることによって、意味が同じ語、つまり言い換え語句に関して類似な文を選択することができた。しかし、言い換え語句ではなく、意味的に上位概念あるいは下位概念、反対概念の語句に関して、調査したい場合がある。例えば、先の例の「取り除くの意」とは逆の「保存するの意」の文を抽出して、両者を比較することによって、用法の特徴がさらにはっきりしてくる。そのためには、同義語辞書と同じように、意味シソーラス辞書を備えればよい。意味シソーラス辞書としては、日本語のものでは「分類語彙表」（国立国語研究所）、「類語新辞典」（角川書店）が知られている。また、外国語のものでは、Roget's Thesaurus (P. M. Roget: "Thesaurus of English Words and Phrase s", London, 1851) をはじめ、多数ある。

【0033】実現方法としては、例えば、図 7～9 に示すような意味マーカの体系を作成し、解析辞書の各見出し語に意味マーカを付与しておく。使用者は、上位あるいは下位、いくつかの階層を同一とみなすかを入力する。照合処理の際には、入力文中の単語と蓄積文中の単語との意味マーカの関係が使用者の入力した階層数以内であれば合致したと判定すればよい。

【0034】実施例 5（請求項 7 に対応）
蓄積されている文の量が多くなると、検索にかかる時間が長くなり、使用感が著しく損なわれる。検索時間を短縮するためには、検索対象文を少なくすればよい。具体的には、蓄積手段によって蓄積されている文を分類し、一まとまりにしておき、使用者が検索範囲を指示できるようにする。特に、言語調査の場合、出典を明らかにすることが望まれるので、出典毎に分類しておけば、検索時間が短くなるだけでなく、言語学的にも意味のある結果を得ることができる。図 10 のブロック図では、複数の蓄積手段 6-1、6-2、6-3 が並列しているように書かれているが、蓄積手段は物理的に分類されている必要はなく、論理的に検索範囲を指定できるのであれば、その形態は問わない。

【0035】本発明は、以上の実施例に限定されるものではなく、入力文を大量に蓄積する手段と複数の言語解析結果を比較・照合する手段と比較・照合結果に基づいて当該結果を選択する手段と比較・照合の実施条件を指定する手段とに加え、言語解析結果を蓄積する手段と第 1 言語文および第 2 言語文の対を入力する手段と、その対を記憶する手段と、その対を大量に蓄積する手段と、その対の一方および両方を言語解析する手段と 2 言語間

の対訳用解析辞書と同義語辞書および意味シソーラス辞書を有することを特徴とし、これによって、（1）入力文と言語的に類似する文を、蓄積された文の中から抽出することによって、入力文中に含まれる語彙の用法を知る機能、（2）入力文と言語的に類似する文の対訳結果を、蓄積された文の中から抽出するの機能を備えた言語解析装置を提供することに本質がある。

【0036】

【発明の効果】請求項 1、2 の発明においては、入力文と言語的に類似する文を蓄積された文の中から抽出することによって、入力文中に含まれる語彙の用法を知ることができる。請求項 3 の発明においては、蓄積文の解析結果に対して照合処理を行なうだけでよく、その都度、蓄積文を構文解析する必要はないため、使用者は、所望の文に類似する文を速やかに得ることができるので、使用感が著しく改善される。請求項 4 の発明においては、所望の文に類似な文の翻訳例を得ることができるので、所望の文を翻訳する際の参考になり、翻訳作業を高精度に行うことができる。請求項 5 の発明においては、第 2 言語文に類似する第 1 言語文を検索することができるので、対訳データを必要としないで翻訳例を得ることができる。請求項 6 の発明においては、意味的に類似あるいは反対の文を検索できるので、用法の特徴をより明確に調査することができる。請求項 7 の発明においては、使用者が検索範囲を指定することができるので、検索時間が短縮され、使用感が著しく改善されるだけでなく、蓄積文を分類して保管できるので、出典を明確にすることが可能になる。

【図面の簡単な説明】

【図 1】 本発明の原理を示すブロック図である。

【図 2】 構文解析による構文木の例である。

【図 3】 類似文抽出処理のフローチャートである。

【図 4】 類似文抽出処理のフローチャートである。

【図 5】 入力文一文に関する照合選択処理のフローチャートである。

【図 6】 対訳（英和）辞書の例である。

【図 7】 意味マーカ体系の例である。

【図 8】 意味マーカ体系の例である。

【図 9】 意味マーカ体系の例である。

【図 10】 複数の蓄積手段を持つ場合の本発明の原理を示すブロック図である。

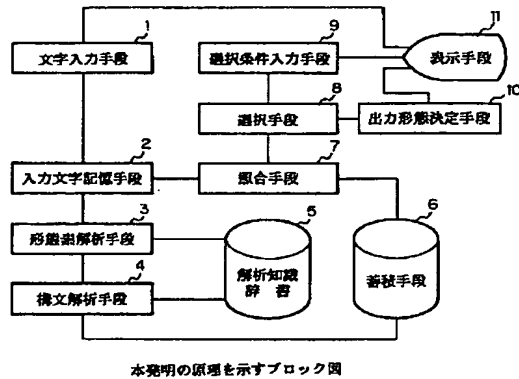
【符号の説明】

1…文字入力手段、2…入力文字記憶手段、3…形態素

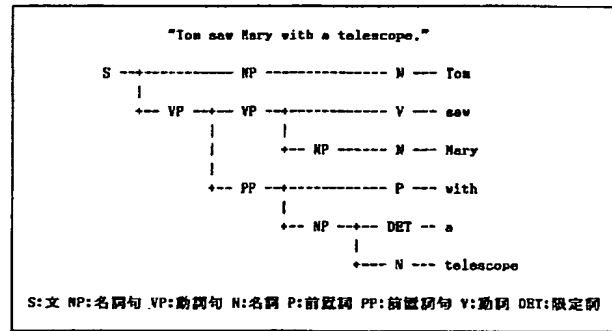
解析手段、4…構文解析手段、5…解析知識辞書、6…蓄積手段、7…照合手段、8…文選択手段、9…選択条

件入力手段、10…出力形態決定手段、11…表示手段。

【図 1】

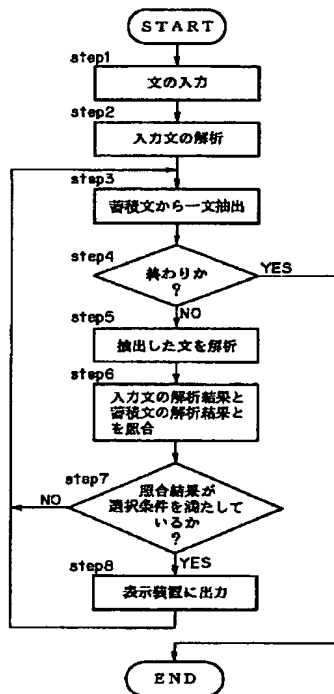


【図 2】



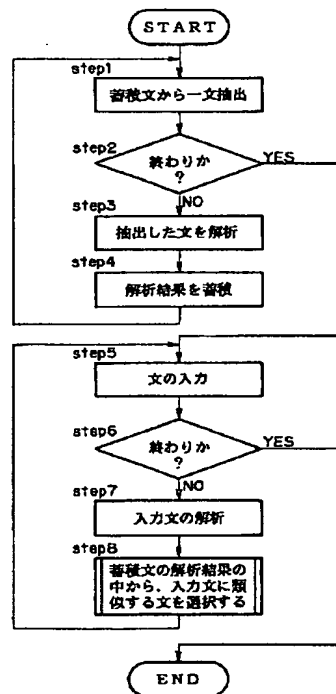
構文木の例

【図 3】



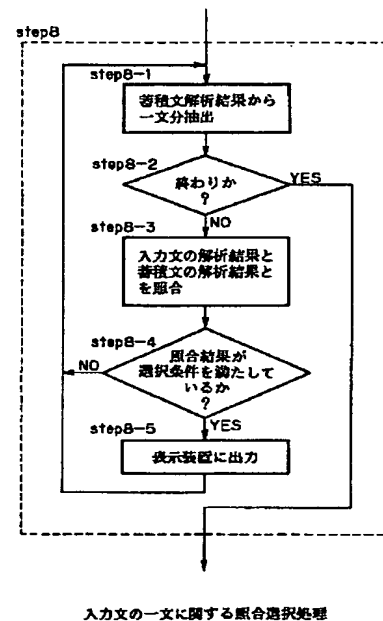
類似文抽出処理のフローチャート

【図 4】



類似文抽出処理のフローチャート

【図 5】

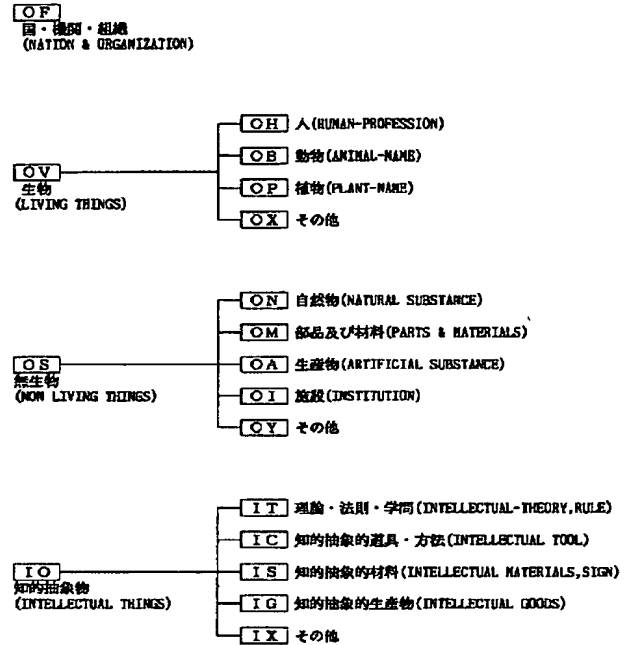


【図6】

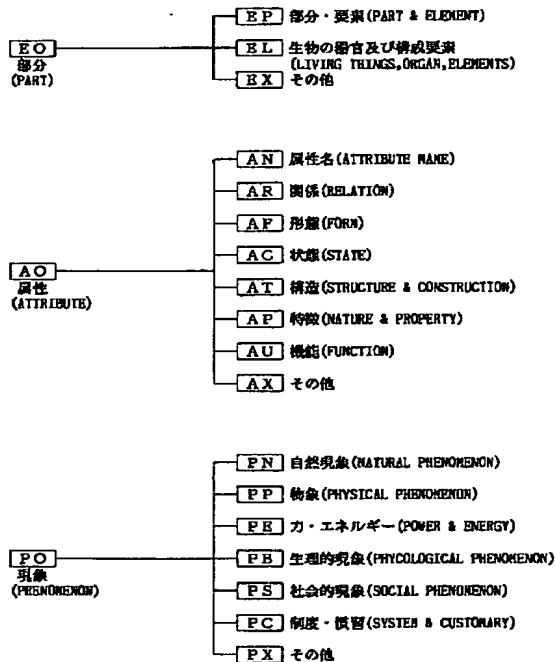
adhesive	〔形〕 粘着性の〔名〕 接着剤; 粘着物
carefully	〔副〕 注意深く; 慎重に; 入念に
excess	〔形〕 超過した; 余分な〔名〕 超過; 過剰; 過度
remove	〔動〕 移す; 除去〔移動〕する; 解任する〔名〕 移動; 移動
shall	〔助〕 ~だろう; ~させる; ~すべきである

対訳(英和)辞書の例

【図7】



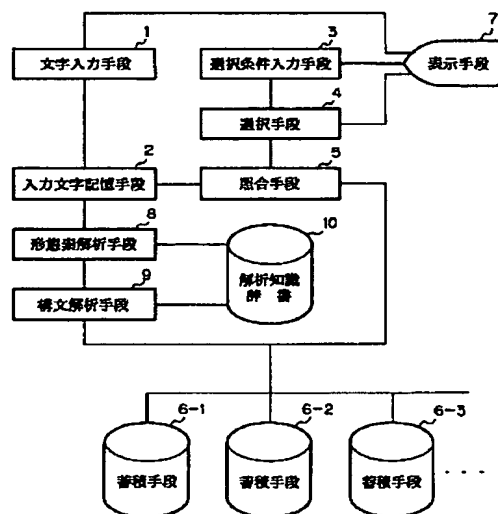
【図8】



意味マーカー体系の例

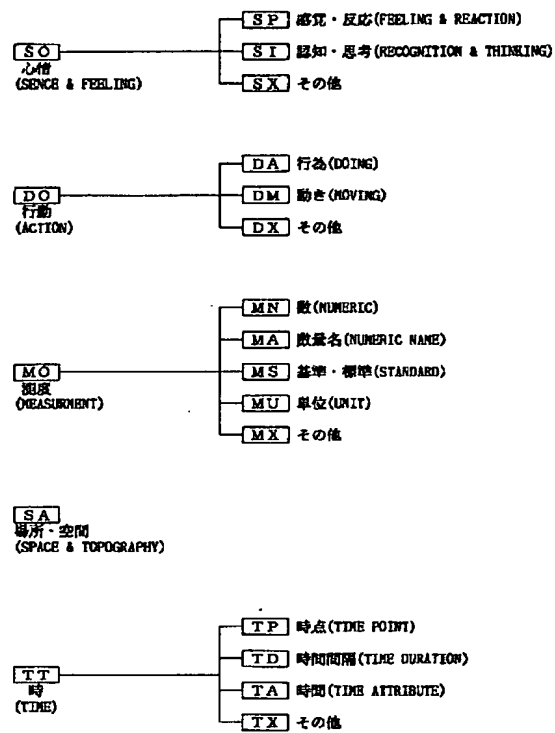
意味マーカー体系の例

【図10】



複数の蓄積手段を持つ場合のブロック図

【図 9】



意味マーカー体系の例

フロントページの続き

(72)発明者 フイリス アンウイル
東京都大田区中馬込 1 丁目 3 番 6 号 株式
会社リコー内

(72)発明者 成田 真澄
東京都大田区中馬込 1 丁目 3 番 6 号 株式
会社リコー内